ComS 573: Machine Learning
Spring 2012

Homework 3
Due Friday, February 3, 2012 in class

Note: Please do not hesitate to contact the instructor or TA if you have difficulty understanding or getting started with solving any of the problems.

1. (20 pts.) Let x have a uniform density

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

   Suppose that n samples $D = \{x_1, \ldots, x_n\}$ are drawn independently according to $p(x|\theta)$. Show that the maximum-likelihood estimate for θ is $\max_k x_k$ – that is, the value of the maximum element in D. Explain in words why you do not need to know the values of the other points.

2. (40 pts.) Let $\vec{x}$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\vec{x}|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

   where $\vec{\theta} = (\theta_1, \ldots, \theta_d)^t$ is an unkown parameter vector, θ_i being the probability that $x_i = 1$. Let $D = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a set of n samples independently drawn according to this probability density.

   (a) If $\vec{s} = (s_1, \ldots, s_d)^t$ is the sum of the n samples ($\vec{s} = \sum_{j=1}^{n} \vec{x}_j$), show that

$$P(D|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

   (b) Show that the maximum-likelihood estimate for $\vec{\theta}$ is

$$\hat{\vec{\theta}} = \frac{\vec{s}}{n}$$

   (c) Assuming a uniform prior distribution for $\vec{\theta}$

$$p(\vec{\theta}) = \begin{cases} 1 & 0 \leq \theta_i \leq 1,\ i = 1, \ldots, d \\ 0 & \text{otherwise,} \end{cases}$$

and using the identity

$$\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!},$$

show that

$$p(\vec{\theta}|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i}.$$

(d) Show that

$$P(\vec{x}|D) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}.$$

(e) If we think of obtaining $P(\vec{x}|D)$ by substituting an estimate $\hat{\vec{\theta}}$ for $\vec{\theta}$ in $P(\vec{x}|\vec{\theta})$, what is the effective Bayesian estimate for $\vec{\theta}$?